ELSEVIER
Available online at www.sciencedirect.com
SciVerse ScienceDirect
Physics Procedia 24 (2012) 176 – 183
Physics Procedia

2012 International Conference on Applied Physics and Industrial Engineering

A Spatial Load Forecasting Method Based on the Theory of Clustering Analysis

Xiao Bai[1], Guo Peng-wei[1], Mu Gang[1], Yan Gan-gui[1],Li Ping[1],Cheng Hong-wei[2], Li Jie-fu[2] and Bai Yang[2]

[1]*School of Electrical Engineering, Northeast Dianli University, Jilin, China*
[2]*Jilin Power Supply Corporation, Jilin, China*

Abstract

The spatial load forecasting method based on time series analysis usually divides predictive zone into cellules, and then forecast the load of each cell based on time series analysis, to predict the whole spatial load. Now there are many prediction models based on time series analysis, but it is hard to realize the minimum of the prediction error for each cell, namely not guarantee the accuracy of results of spatial load forecasting. Thus, the spatial load forecasting method based on the theory of clustering analysis is proposed there. Firstly this method analyzes the results and the relative prediction error of each cell with different load forecasting model. Second, according to the best forecast models, cluster all cellules, which cellules in the same cluster use the same prediction model to forecast the load of cellules of the target year. The results of a practical example show that our method is correct and effective.

*Keywords:*power system; spatial load forecasting (SLF); the theory of clustering analysis

1. Introduction

Spatial load forecasting (SLF) is necessary in distribution planning, and the accuracy of the calculation directly affect needed capacity and distribution of the distribution equipment. Spatial load forecasting (SLF) [1,2], also called for small area load forecasting, is the power load forecasting of the city in the space-time distribution. Currently, the main SLF methods are direct observation method, multivariate method, time series analysis method and land usage method.

Direct observation method mainly relies on the planning personnel experience and subjective judgment to decide the size and distribution of the load, and it lacks necessary scientificity in a certain extent. Multivariate method is based on each district's annual peak load of historical data and other

 doi:10.1016/j.phpro.2012.02.027

variables as a basis for predicting peak load of the target year, but quality of required data is relatively high [3]. It has been phasing out in the 80's.

Land usage method predicts urban land-use types, use intensity, land values and their geographical distribution and size composition with the means of analyzing the characteristics of urban land use and development laws, and converts land-use into spatial load on this basis [2,3,4,5]. This method is a top-down forecasting method, and it can meet the long-term load forecasting, but it's more suitable for a perfect market economy, the domestic need to be studied [1,2,6,7].

Time series analysis method divides estimation range into cellular regions, and then uses the cellular historical load data to establish an appropriate forecasting model to predict the future value on the load. Time series analysis method is a bottom-up forecasting method, with the advantage of only needing less the historical load data. Commonly used forecasting method are index smoothing forecasting method (ISFM), grey forecasting method (GFM), linear regression forecasting method (LRFM) and so on.

There are a lot of spatial load forecasting modes based on time series analysis, but not any single model can realize the minimum of the forecasting error for each cell. Towards this problem, this article proposes the SLF method based on the theory of clustering analysis. This method clusters all cellules by analysising predictive results and the relative error of each cell with different load forecasting models, which cellules in the same cluster forecast the load of cellules of the target year using the same prediction model. Changyi District in a certain city as an example verifies the validity of the method.

2. The basic principle of clustering

Clustering is a physical or abstract process of grouping and collecting of objects [11]. The most commonly used clustering methods include partitioning method, hierarchical clustering method, density-based method, grid-based method, model-based method and so on [12].

This article clusters cellules using partitioning method. The main idea of this method is that a data object with n sets of data, according to the similarity between objects by iterative data, are divided into k blocks, which each divided block is a cluster, and makes the objective function is minimum.

3. the Spatial Load Forecasting Method Based on the Theory of Clustering Analysis

3.1 The basic principle of SLF method based on clustering

SLF method based on clustering (SLFMBC) can effectively reduce forecasting error of the load of cell which predicted by using a single forecasting model, and improve the predictive accuracy.

The basic principle of SLF method based on clustering is to analyze the results of prediction and the relative forecasting error of cellular load using different models to predict, according to the theory of cluster analysis, to divide cellules which using the same forecasting model have the best forecasting effects into a cluster, cellules in the same cluster using the same prediction model to forecast the load of cellules of the target year.

3.2 The basic steps of SLF method based on clustering

1) Cell division and historical load data collection

Cell division is a necessary step in SLF, which aims to predict the location of load growth, provide spatial information for the planning of distribution network [10]. In geographic information systems the prediction range will be divided into different miniregions, each miniregion is called cell.

Collect the historical load data of each cell and use of data mining technology on historical load data of cellules preprocessing singular data.

2) Prediction using different load forecasting models

Using different forecasting models forecast load of each cell, take the above-mentioned three kinds of SLF models time based on time series analysis as an example. If there are n cellules, the value of cell's load, using the i kinds of prediction model predict the m kinds of cell, is $F_{im}(i = 1,2,3)$.

3) Clustering of cellules

The basic idea of cellular clustering is to forecast different cell by using different forecasting models, and cellules the relative prediction error which using the same forecasting model gets is the minimum are divided into a cluster.

Assumption is known that the historical load data of cell m from 1 to t years, $F_{im}(t)$ that is the predictive value of load of the t kinds of years using the forecasting models i to predict the load of cell m, $Z_m(t)$ that is the actual value of the t kinds of years of cell m, and then $E_{im}(t)$, the forecasting relative error of the t kinds of years using the forecasting model i to predict the load of cell m, is:

$$E_{im}(t) = \frac{F_{im}(t) - Z_{im}(t)}{Z_{im}(t)} \tag{1}$$

If $E_{jm}(t) = \min_{1 \le i \le 3} E_{im}(t)$, cell m is divided into cluster j.

4) Load forecasting of cellules in different clusters

Forecast the load of cellules that have clustered. Predict the load of cellules of the target year in cluster i using the forecasting model i.

4. Example Analysis

Changyi District in a certain city as an example, the largest historical load data is known to every six months in 2004 -2008 years. Predict the load in this region using the SLF method based on the theory of clustering analysis.

4.1 Cell division and historical load data collection

In geographic information systems draw the power supply range of Changyi Power Supply Bureau of a certain city, as shown in figure 1.

Figure 1. The power supply range of Changyi Power Supply Bureau

Divide the power supply range into cellules according to the power supply area of feeders. The power supply area of each feeder is a cell, as shown in figure 2.

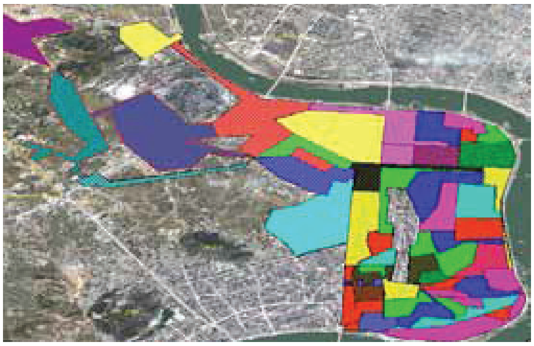

Figure 2. The power supply area of cellules of Changyi District

Collect the largest historical load data of every six months in 2004-2008 years, and use of data mining technology on historical load data of cellules preprocessing singular data.

4.2 Prediction of the first half of 2008 peak load with different forecasting models

According to the largest cellular historical load data of six months, using GFM, ISFM, and LRFM, predict of the first half of 2008 peak load, and calculate absolute average error (AAE), absolute average relative error (AARE) and root-mean-square (RMS), the results as shown in Table 1,2.

TABLE I. THE PREDICTIVE VALUE OF THE FIRST HALF OF 2008 LOAD OF CHANGYI DISTRICT WITH DIFFERENT FORECASTING MODELS/WM

Cell Name	Actual Value (MW)	Predictive Value (MW)		
		GFM	*ISFM*	*LRFM*
Bachang	0.575	0.5508	0.3459	0.4116
Changsha	1.5	1.2104	1.3769	1.1929
Dandong	1.785	3.3648	3.6003	3.4054
Dongbao	1.02	1.1176	1.0074	1.1743
Dongkang	3.18	2.6576	2.8779	2.6243
Dongqing	5.32	4.9121	5.0536	4.7114
Ershui	3.232	2.9041	2.8087	2.928
Fujia	0.568	0.4141	0.3993	0.413
Fuzhuan	3.504	3.3035	3.3611	3.3154
Gongre	4.215	4.263	3.996	4.9271
Guomao	2.05	2.1558	2.0372	2.1843
Hada	1.92	2.131	1.6121	1.4057
Heping	3.568	2.7431	2.8571	2.632
Hongbo	2.304	2.4528	2.1829	2.526
Huadong	1.395	1.2549	0.8458	1.031
Huazaiyi	1.665	2.9962	2.2762	3.1714
Jianshe	2.113	2.1695	2.3385	2.2007
Jianghua	2.534	2.2649	2.2618	2.2764
Jiangwan	4.219	3.3461	3.3654	3.5094
Jiutai	1.42	1.3104	1.294	1.2757
Kangzhuangjia	0.348	0.3303	0.3291	0.32
Kangzhuangyi	3.645	4.1889	3.7601	4.3254
Lianjijia	4.551	5.4223	5.6739	5.5584
Lianjiyi	0.725	0.7561	0.6729	0.7699
Liaodongyi	0.993	0.8971	0.9077	0.9189
Linshan	1.288	1.2632	1.3955	1.1486
Nenjiang	2.093	2.1301	2.1635	2.0269
Nongan	1.64	1.5374	1.6375	1.5386
Sanshui	2.11	1.8343	2.0976	1.9414
Shahe	4.632	5.3737	4.1302	4.7637

Shangmao	0.52	0.4898	0.4415	0.4957
Shanghaijia	3.39	2.8796	2.9944	2.8643
Shanghaiyi	2.74	2.5164	2.6258	2.47
Sichuan	2.869	0.9968	0.8655	0.859
Songhua	1.961	1.6544	1.9053	1.537
Tianjin	4.267	3.9572	3.8411	4.12
Wanda	1.962	2.1615	1.8838	2.0677
Weichang	1.066	0.961	0.9834	0.9677
Xinchun	0.816	0.7494	0.7128	0.7837
Xinyu	3.369	4.9195	3.7128	4.5113
Xuzhou	3.64	3.4755	3.1868	3.4457
Youdian	2.84	2.7516	2.7472	2.7214
Zhizai	0.94	1.3223	0.4449	1.2971
Zhongxing	1.4	1.0915	1.2642	1.0871
Chongqing	1.304	1.6184	1.4124	1.4344
Zhuanchang	2.355	2.7923	2.8176	2.7766

TABLE II. AAE, AARE AND RMS OF THE FIRST HALF OF 2008 LOAD OF CHANGYI DISTRICT WITH DIFFERENT FORECASTING MODELS

Forecasting Model	AAE (MW)	AARE (%)	RMS (MW)
GFM	0.3817	16.60	0.3925
ISFM	0.3263	15.49	0.3741
LRFM	0.4001	18.23	0.3894

4.3 Clustering of cellules

Analyze the forecasting results and the relative error of the first half of 2008 with different forecasting models, and then according to the principle of forecasting error minimum, divide all cellules of Changyi District into grey cluster (GC), index smoothing cluster (ISC) and regression analysis clusters (RAC), clustering results as shown in Table 3; in geographic information systems the distribution of clusters as shown in Figure 3.

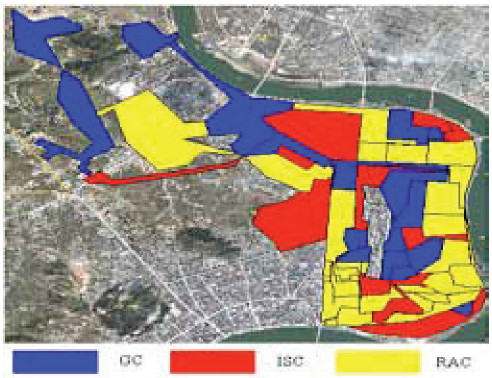

Figure 3. The distribution of clusters

TABLE III. CLUSTERING RESULTS

Cluster Name	Cell Name
GC	Bachang, Dangdong, Fujia, Gongre, Hada feeder, Huadong, Jianshe, Jiutai, Kangzhuangjia, Lianjijia, Lianjiyi, Linshan, Nenjiang, Sichuan, Xuzhou, Youdian
ISC	Changsha, Dongbao, Dongkang, Dongqing, Fuzhuan, Guomao, Heping, Hongbo, Huazaiyi, Kangzhuangyi, Nongan, Sanshui, Shanghaijia, Shanghaiyi Songhua, Wanda, Weichang, Xinyu, Zhongxing, Chongqing
RAC	Ershui, jianghua, Jiangwan, Liaodongyi, Shahe, Shangmao, Tianjin, Xinchun, Zhizai, Zhuanchang

4.4 Prediction of the second half of 2008 peak load

Forecast the second half of 2008 peak load in different clusters with the corresponding forecasting model, forecasting results as shown in Table 4; calculate AAE, AARE and RSM, the results as shown in Table 5.

TABLE IV. The Predictive value of the second half of 2008 load of Changyi District with different forecasting models/WM

Cell Name	Actual value (MW)	Predictive Value (MW)			
		GFM	*ISFM*	*LRFM*	*SLFMBC*
Bachang	0.568	0.6589	0.615	0.5933	0.5933
Changsha	1.38	1.4744	1.5004	1.3329	1.3329
Dandong	1.963	2.3794	0.6854	2.3939	2.3794
Dongbao	0.84	0.9989	0.9543	1.0657	0.9543
Dongkang	2.87	2.915	3.1687	2.8029	2.915
Dongqing	4.29	5.3098	5.2214	5.5304	5.2214
Ershui	2.84	3.0884	3.1867	3.0469	3.0469
Fujia	0.568	0.5362	0.4195	0.5113	0.5362
Fuzhuan	3.736	3.3225	3.4919	3.2903	3.4919
Gongre	3.36	3.5697	3.7807	4.3114	3.5697
Guomao	2.02	2.1513	2.0902	2.1771	2.0902
Hada	1.752	1.8195	2.0374	1.6046	1.8195
Heping	3.184	3.6086	3.2263	3.0336	3.2263
Hongbo	2.226	2.3226	2.3173	2.4934	2.3173
Huadong	1.59	1.6386	1.412	1.3869	1.6386
Huazaiyi	1.56	2.2085	0.9934	2.31	0.9934
Jianshe	2.042	1.9428	2.0253	1.9276	2.0253
Jianghua	2.114	2.4809	2.5834	2.46	2.46
Jiangwan	3.607	3.8059	4.0421	3.7203	3.7203
Jiutai	1.33	1.3684	1.4661	1.37	1.3684
Kangzhuangjia	0.335	0.3339	0.3491	0.339	0.3339
Kangzhuangyi	2.753	3.7823	3.3164	3.8844	3.3164
Lianjijia	6.021	4.9006	4.1773	4.8537	4.9006
Lianjiyi	0.67	0.7138	0.7157	0.7481	0.7138
Liaodongyi	0.956	0.933	1.0166	0.9347	0.9347
Linshan	1.368	1.3405	1.397	1.2766	1.3405
Nenjiang	2.235	2.0926	2.0641	2.1001	2.0926
Nongan	1.65	1.5815	1.6924	1.5836	1.6924
Sanshui	2.11	1.8872	2.1049	1.8286	2.1049
Shahe	5.064	5.5664	4.3231	5.2571	5.2571
Shangmao	0.61	0.4893	0.5086	0.5201	0.5201
Shanghaijia	5.67	3.0964	3.4424	3.0414	3.4424
Shanghaiyi	3.94	2.767	2.8353	2.6786	2.8353
Sichuan	3.978	—	4.5736	2.1256	4.5736
Songhua	2.123	1.9241	1.9562	1.7026	1.9562
Tianjin	4.206	3.934	4.1958	4.05	4.1958
Wanda	2.236	2.1984	1.8298	2.2016	2.2016
Weichang	0.92	0.9738	1.0758	0.9809	0.9738
Xinchun	0.627	0.7588	0.7439	0.8171	0.7439
Xinyu	3.874	4.3038	3.6259	4.2991	3.6259
Xuzhou	3.448	3.5954	3.6389	3.6091	3.5954
Youdian	2.44	2.79	2.8845	2.7967	2.79
Zhizai	0.94	1.0257	0.8123	0.9057	0.9057
Zhongxing	1.58	1.0849	1.3797	1.0771	1.3797
Chongqing	1.278	1.5024	1.4037	1.5577	1.4037
Zhuanchang	2.985	2.4517	2.0826	2.5377	2.5377

TABLE V. AAE, AARE and RMS of the first half of 2008 load of Changyi District with different forecasting models

Forecasting Model	AAE (MW)	AARE (%)	RMS (MW)
GFM	0.4095	12.63	0.3861
ISFM	0.3702	14.92	0.3816
LRFM	0.3947	14.81	0.3881
SLFMBC	0.2592	3.29	0.391

4.5 Results and Analysis

Analyze the forecasting results and the relative error of the first half of 2008 with different forecasting models in table 4, and comparing with results in table 1, the number of cellules the most effective forecasting model of the first half of 2008 and the second half of 2008 is coincident are 36, accounting for 78.26%, and the number of cellules the most effective forecasting model of the first half of 2008 and the second half of 2008 is not coincident are 10, accounting for 21.74%. The number of cellules the most effective forecasting model of the first half of 2008 and the second half of 2008 is ISFM, GFM and LRFM are 13, 16 and 7, accounting for 28.26%, 34.78% and 15.22%. The comparison between models is shown in Figure 4.

From the table 5, comparing with the single SLF model, the SLF method based on the theory of clustering analysis reduces the error of SLF, and improve the accuracy of the predictions.

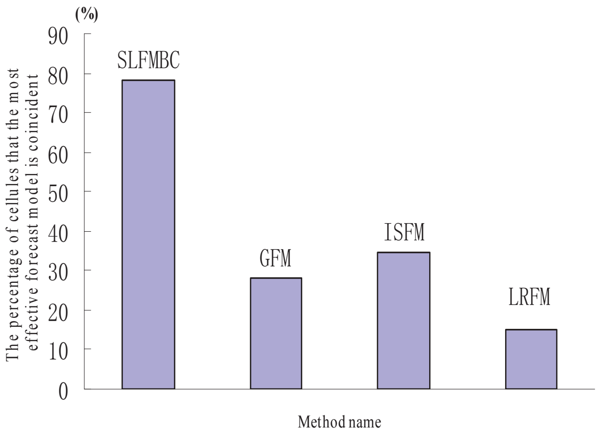

Figure 4. Comparison of various methods

Conclusion

According to the problem that a single forecasting model cannot guarantee the accuracy of spatial load forecasting results which generated in total forecasting area, this paper presents a spatial load forecasting method based on the theory of clustering analysis. This method is based on traditional time series analysis method. Fristly, analyze the results and the relative error of prediction of each cell with different load forecasting model, and then according to the best forecast model, cluster cellules, which in the same cluster use the same prediction model to forecast the load of cellules of the target year for cellules. The results indicate the spatial load forecasting method based on the theory of clustering analysis can reduce the prediction error, improve forecasting accuracy, and also have a certain practicality.

References

[1] H. L. Willis, and E. D. James, "Spatial electric load forecasting," Proceeding of the IEEE, pp. 252-253, Feb. 1983.

[2] H. L. Willis, and H. Tram, "Cluster based V.A.I. method for distribution load forecasting," IEEE Transactions on Power Apparatus and Systems, pp. 2677-2684, Aug. 1983.

[3] WANG Tian-hua, WANG Ping-yang, YUE Zhong-bin, “Spatial load forecasting for distribution planning based on GIS platform," Power System Technology, vol. 25, May. 1999, pp. 42-47.

[4] YANG Li-xi, WANG Jin-feng, CHEN Gen-yong, WANG Jia-yao, “Load spatial distribution forecasting model on celluar automata theory," Proceedings of the CSEE, vol. 27, Apr. 2004, pp. 15-20.

[5] WANG Tian-hua, WANG Ping-yang, FAN Ming-tian, “Application of genetic algorithm, fuzzy logic and transshipment model to spatial load forecasting for distribution planning,” Power System Technology, vol. 23, Jan. 1999, pp. 24-28.

[6] CHENG Hao-zhong, Power System Planning Power, Beijing: China Electric Power Press, 2008, pp. 47-48.

[7] KANG Chong-qing, XIA Qing, LIU Mei, Power System Forecasting, Beijing: China Electric Power Press, 2007, pp. 68-73.

[8] W. Charytoniuk, M. S. Chen, and Van Olinda P. Nonparametric, "Regression based short-term load forecasting," IEEE Transaction on Power System, pp. 725-730, Mar. 1998.

[9] ZHOU Ping, YANG Lan, ZHOU Jia-qi, "The new method of Electric Power System Load Gray Forecast," Proceedings of the CSU-EPSA, vol. 10, Mar. 1998, pp. 45-50.

[10] MA Xiao-dong, ZHUAN Min-hui, SHI Lin-jun, "Application of Spacial Load Forecasting in Power Grid Planning of Suzhou Industrial Park,", Jiangsu Electrical Engineering, vol. 25, Mar. 2006，pp. 24-26.

[11] ZENG Zhi-xiong, "Effective hybrid clustering algorithm based on partition and hierarchy," Journal of Computer Applications, vol. 27, Jul. 2007，pp. 33-36.

[12] JIA Hui-min, HE Guang-yu, "Load Forecasting by Multi-Hierarchy Clustering Combining Hierarchy Clustering with Approaching Algorithm in Two Directions," Power System Technology, vol. 23, Dec. 2007, pp. 33-36.